

ARI Research Note 96-70

An Empirical Assessment of Coaching and Practice Effects on Three Army Tests of Spatial Aptitude

Henry H. Busciglio and Dale R. Palmer
U.S. Army Research Institute

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

July 1996

19970110 001



1996 QUALITY INSPECTED 4

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

**EDGAR M. JOHNSON
Director**

Technical review by

Elizabeth J. Brady
Martha Teplitzky

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, July		2. REPORT TYPE Final		3. DATES COVERED (from... to) March 1991-June 1992	
4. TITLE AND SUBTITLE An Empirical Assessment of Coaching and Practice Effects on Three Army Tests of Spatial Aptitude				5a. CONTRACT OR GRANT NUMBER	
				5b. PROGRAM ELEMENT NUMBER 0603007A	
6. AUTHOR(S) Henry H. Busciglio and Dale R. Palmer				5c. PROJECT NUMBER A792	
				5d. TASK NUMBER 1222	
				5e. WORK UNIT NUMBER H01	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Research Note 96-70	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT (<i>Maximum 200 words</i>): Three Army tests of spatial aptitude--Assembling Objects, Figural Reasoning, and Orientation--were included in the Enhanced Computer Administered Testing (ECAT) project. ECAT was a joint service effort to evaluate measures for possible addition to future versions of the Armed Services Vocational Aptitude Battery (ASVAB). Because both practice and coaching effects might threaten the long-term validity of these tests, U.S. Army Research Institute for the Behavioral and Social Sciences researchers assessed their susceptibility to such effects. Overall, we found coaching and practice effects that are comparable to those obtained in previous research using spatial tests. The Orientation test was the only measure for which specific coaching led to significantly larger effect sizes than did practice alone. These results have ramifications for future research and development, such as exploring ways to: (1) lower the susceptibility of the Orientation test to coaching, and (2) reduce practice effects on the Assembling Objects and Figural Reasoning tests. We hope these activities help to ensure the long-term validity of these Army tests.					
15. SUBJECT TERMS Spatial tests Coaching and practice effects ASVAB					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT Unlimited	20. NUMBER OF PAGES 33	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified			

ACKNOWLEDGMENTS

We would like to express our appreciation to the following individuals for their valuable help:

- Dr. Clinton B. Walker, for many important contributions.
- Personnel of the 120th AG Bn at Fort Jackson and of TRADOC for arranging the data collections.
- The American Institutes for Research and RGI, Inc. for helping to create test materials, collect data, and create the data files.

Special thanks also to Dr. Michael Rumsey, Selection and Assignment Research Unit Chief, and to Drs. Elizabeth Brady and Martha Teplitzky for their astute and constructive comments on earlier drafts of this paper.

AN EMPIRICAL ASSESSMENT OF COACHING AND PRACTICE EFFECTS ON THREE ARMY TESTS OF SPATIAL APTITUDE

EXECUTIVE SUMMARY

Research Requirement:

The purpose of this research was to empirically assess the impact of practice and coaching on three of the Army's Project A tests of spatial aptitude. These measures (Assembling Objects, Figural Reasoning, and Orientation) were included in the Enhanced Computer Administered Testing (ECAT) project, a joint service effort to evaluate measures for possible addition to the Armed Services Vocational Aptitude Battery (ASVAB). Because practice and coaching effects might threaten the long-term validity of these tests, we wanted to determine their susceptibility to such effects and, if possible, gain insights into the most feasible countermeasures.

Procedure:

U.S. Army Research Institute for the Behavioral and Social Sciences researchers studied the items on the three tests for generalities or commonalities that might serve as useful hints for teaching individuals how to do better on the tests (without any real learning or improvement in spatial skills). These hints were organized into a set of coaching instructions. After several pilot efforts we decided that the best mode of teaching would be to have the instructions audio-taped and to have subjects respond to printed instructional materials while listening to the tapes. To assess the vulnerability of the measures to a more general type of coaching, we examined several publications and developed a brief handout containing such "hints" as guessing, time management, and other topics pertinent to multiple-choice tests.

Findings:

We tested a group of 1,915 new Army recruits at Fort Jackson SC, in June of 1992. Subjects were assigned to groups that received either specific or general coaching, or practice alone, on one of the three tests. Overall, we found that the tests are subject to coaching and practice effects of a size about equal to the effects obtained in previous research using spatial tests. The Orientation test was the only measure for which specific coaching led to significantly larger effect sizes than did practice alone. On all three tests, general coaching was no more effective than practice. Posttest gain scores were significantly related to self-reported use of coaching and expectations of

score improvements due to coaching. These responses also indicated subjects' belief the coaching like ours could be expected if the tests were made operational.

Utilization of Findings:

Our findings show that only the Orientation test is especially vulnerable to a "quick-and-easy" coaching strategy. We would therefore recommend that certain content changes be made to lessen its susceptibility to coaching. For the other measures, Assembling Objects and Figural Reasoning, effective coaching would involve much more extensive, time-consuming procedures. However, practice effects on these tests might be large enough to warrant countermeasures such as giving all examinees more practice items to complete immediately before the test itself, or including, in future ASVAB orientation materials, incentives, and opportunities to practice before the test session.

AN EMPIRICAL ASSESSMENT OF COACHING AND PRACTICE EFFECTS ON THREE ARMY TESTS OF SPATIAL APTITUDE

CONTENTS

	Page
Introduction	1
Project A Spatial Tests	1
Previous Research on Coaching and Practice Effects on Spatial Tests	1
Popular Coaching Books and Publications	3
Method	4
Development of Specific Coaching Strategies and Materials	4
Development of General Coaching Strategy	5
Subjects	5
Testing Schedules	6
Posttest Questionnaire	6
Results	7
Within-Subjects Analyses	7
Between-Subjects Analyses	8
Analysis of Posttest Questionnaire Responses	12
Discussion	14
References	19
APPENDIX A - Statistics for Number Correct Scores	A-1
APPENDIX B - Statistics for Speed Scores	B-1
APPENDIX C - Statistics for Accuracy Scores	C-1

LIST OF TABLES

Table 1. Effect Sizes and Significance Tests for Within-Subjects Analysis of Coaching and Practice Effects	8
Table 2. Between-Subjects Analysis of Effects of Coaching After Practice and Practice Only	10
Table 3. Between-Subjects Analysis of Effects of Coaching Before Practice and Practice Only	11
Table 4. Relationship of Gain Scores to Number of Coaching Strategies Remembered	13

AN EMPIRICAL ASSESSMENT OF COACHING AND PRACTICE EFFECTS ON THREE
ARMY TESTS OF SPATIAL APTITUDE

CONTENTS (continued)

	Page
Table 5. Relationship of Gain Scores to Self-Reported Use of Coaching Strategies	14
Table 6. Relationship of Gain Scores to Recipients' Perceptions of the Improvement of Test Scores Due to Coaching	15
Table 7. Perceived Likelihood of Coaching Like Ours if Spatial Testing Became Operational	16

An Empirical Assessment of Coaching and Practice Effects on Three Army Tests of Spatial Aptitude

Introduction

Project A Spatial Tests

Under the U.S. Army's Project A (e.g., Campbell & Zook, 1991), 16 new aptitude tests were developed and evaluated as measures to supplement the Armed Services Vocational Aptitude Battery (ASVAB), the Army's operational selection instrument. Based on good showings in Project A, several spatial tests - Assembling Objects, Figural Reasoning, and Orientation - are now being considered for addition to the ASVAB. It is therefore important that these tests remain valid incremental predictors (over and above ASVAB) of various job performance criteria. One important potential threat to the long-term validity of these tests is the confounding of true spatial ability with differential practice and coaching effects. This research was meant to investigate this threat.

Previous Research on Coaching and Practice Effects on Spatial Tests

We began this investigation by surveying the previous literature on the effects of practice and/or coaching on spatial test scores. Although the research literature is not extensive and specific principles are rare, several sources do suggest, in general, that practice and coaching can affect spatial scores in nontrivial ways.

In an early review of practice/training effects upon perceptual judgements, Gibson (1953) reported studies that found that practice and/or training significantly improved such spatially oriented skills as estimating the linear extent, area, and angles of various geometric figures. Goldstein and Chance (1965) found substantial practice effects on a set of items taken from the Embedded Figures Test and two other measures of field dependence. Brinkmann (1966) investigated the effects of programmed instruction as a technique for improving "spatial visualization." The author found that subjects receiving the programmed instruction scored significantly higher on the spatial tests than those in the control group. Saunderson (1973) also found that an experimental group given specific training on spatial tasks scored significantly higher on later spatial ability tests. Sherman (1974) obtained a significant practice effect on a measure of field articulation called the Rod-and-Frame Test.

Conner, Schackman, and Serbin (1978) noted both practice and training effects on a children's form of the Embedded Figures

Test. In this study, training consisted of brief visual presentations and explanations of items like those used on the test itself; practice effects were noted as an increase in the posttest scores of an untrained control group. McGee (1978) found that training, in the form of "a one-hour lecture on spatial abilities," significantly improved scores on a five-item form of the Mental Rotation Test. Kyllonen, Lohman, and Snow (1984) noted that training strategies and performance feedback were both effective in increasing scores on spatial tests.

Stericker and LeVesconte (1982) found that a group of experimental subjects, after being exposed to three hours of practice and training, did significantly better on four standard tests of visual-spatial skill. The practice and training received by the experimental group involved reviewing the solutions for various example items and items missed on a pretest. Physical models of some spatial test problems were available for the subjects to rotate and compare visually to each of four possible answers.

Stericker and LeVesconte (1982) also found that their coaching effects were transferable. Specifically, training significantly improved posttest performance on three "trained" tests and one "untrained" test. This finding suggests that some training effects, on certain spatial tests, generalize beyond the immediate training situation and serve to increase scores on other tests of spatial ability as well. On the other hand, Gagnon (1985) found that a five-hour training session on two video games did not, in general, lead to significant differences between the mean scores of the training and control groups on four measures of spatial ability. This suggests that spatial training does not necessarily generalize.

In a recent meta-analysis based on nine different samples, Baenninger and Newcombe (1989) found that "specific" training (i.e., training on a single spatial measure) produced significant increases in spatial scores. However, the same authors found that "short" training- i.e., single administrations or brief administrations over a period of less than three weeks [usually the case with specific training] - produced effect sizes that were not significantly different from those of practice-only. The authors concluded that "brief training fulfills the same function as practice. That is, it enhances test-specific spatial ability but not necessarily general spatial ability" (p. 339).

Peterson (1987), in his report on the development of the Project A tests, included data on the three spatial measures involved in the present study.¹ A sample of individuals was

¹An early, 40-item version of the Assembling Objects test was used.

retested two weeks after taking the measures for the first time. The results are summarized below:

Test	Time 1		Time 2		Effect Size ²
	Mean	SD	Mean	SD	
Assembling Objects	25.68	9.13	28.23	8.84	0.279
Orientation	11.64	5.99	12.31	6.12	0.112
Figural Reasoning	20.35	5.03	21.15	5.49	0.159

Silva and Busciglio (1993, p. 6) have recently summarized the research on practice and coaching effects on spatial test scores and have drawn the following overall conclusions:

The overall range of effect sizes found in the studies of practice effects on spatial test scores is .06 to 1.60 of a standard deviation.³ Additionally, the studies provide some support for the following influences upon effect sizes: 1. inter-test period - shorter inter-test periods generally lead to larger effect sizes; 2. type of score - latency, or reaction time, or other types of speed scores lead to larger effect sizes than do accuracy (i.e., proportion correct) scores; and 3. type of retest - effect sizes are apparently higher when the same form is used on the retest, instead of a different form.

Effect sizes [for coaching] range from .29 to 1.26. One consistent pattern here seems to be the larger effect sizes for specific, as opposed to general, coaching. Specific coaching may be described as any type of instruction (usually accompanied by practice) on items that are identical, or very similar to, the actual items on the test. In contrast, general coaching is less similar to the content of the test and may range from playing video-games, to lectures, to brief handouts on doing better on multiple-choice tests.

Popular Coaching Books and Publications

We continued our review by surveying actual "coaching" books that are easily accessible to the public. Two examples are: Up the IQ! by Paul I. Jacobs (1977) and Know Your Own I.Q. by H. J. Eysenck (1962). Jacobs' book presents a list of 12 principles or rules for solving test items that require the subject to

²Like others reported in this paper, these effect sizes are equal to the difference between Time 2 and Time 1 means, divided by the Time 1 standard deviation.

³In most studies, "practice" means having taken the same test previously; in some cases, it means something else, such as taking an alternate form or having multiple exposures to the same test.

determine patterns of change or constancy within groups of figures. Examples and practice items are given to facilitate learning. The Army's Figural Reasoning test, one of the measures considered in this research, contains items that are very similar to those coached in Jacobs' book. Eysenck's book also discusses this type of item and gives readers the correct answers and strategies to obtain them.⁴

Based upon the literature reviewed above, there appears to be cause for concern about the long-term validity of the Project A spatial tests. First, spatial test scores in general seem to be susceptible to coaching and/or practice effects, at least under some conditions. Second, coaching aids are readily available for this purpose. The possibility thus exists that future scores on the Project A Spatial tests may be invalid measures of true spatial ability due to the confounding influence of differential coaching and/or practice experience. However, since coaching involves training on specific test items (Anastasi, 1982), we concluded that the results of our literature review may not generalize to the Project A measures and that more focused research was called for.

Method

Development of Specific Coaching Strategies and Materials

The first step in the present research was to develop specific coaching strategies for the tests. Basically, we tried to create brief clues, or "hints" designed to make the items quicker and easier to solve. The coaching media were printed workbooks showing the strategies with step-by-step examples. Examinees' study of these workbooks was guided by audio-taped instructions.⁵ The following are brief descriptions of the tests.⁶

Assembling Objects. This test was designed to measure a construct called "Spatial Visualization - Rotation," defined as the ability to "mentally manipulate components of two- and three-dimensional figures into other arrangements" (Campbell & Zook,

⁴There are also, of course, many examples of publications meant to improve candidates' scores on the Armed Services Vocational Aptitude Battery (e.g. Barron's, 1989).

⁵The anticipated mode of presentation of our coaching strategies changed several times. We at first tried live presentations with overhead slides. Then, in an attempt to standardize and simplify administration, we tried to video-tape the coaching. Finally, when we realized that our videos were not visually clear enough, we decided on the current mode.

⁶Due to their sensitive nature, all experimental materials, such as the scripts and workbooks used for specific coaching, are subject to limited distribution. For more information, please contact Dr. Michael Rumsey, Selection and Assignment Research Unit Chief.

1991, p. 19). It contains 36 items with an 18-minute time limit. The subject's task involves figuring out how an object will look when its parts are put together. The subject must choose from four possible answers.

There are two types of problems in the test. On Part One, the items show a number of separate pieces, each labeled at one or more places with a small letter (a, b, c). By mentally matching corresponding letters on different pieces, the subject can "see" how they should be connected when the object is put together correctly. On the second part of the test, the pieces of each object are not labeled. Instead, they fit together like parts of a puzzle.

Figural Reasoning. This test measures "Induction," or "the ability to generate hypotheses about principles governing relationships among several objects" (Campbell & Zook, 1991, p. 21). It contains 30 items with a 12-minute time limit. Subjects are presented a series of four figures. The task is to discover the pattern or relationship among the figures and then to select, from five possible answers, the figure that would appear next in the series.

Orientation. This instrument measures "Spatial Orientation," defined as "the ability to maintain one's bearings with respect to points on a compass and to maintain location relative to landmarks" (Campbell & Zook, 1991, p.20). It contains 24 items with a 10-minute time limit. Each item shows a picture within a circular or rectangular frame. The bottom of the frame has a circle with a dot inside it. The picture or scene is not in an upright position, but is described as fixed in the position. The task is to mentally rotate the frame so the bottom of the frame is positioned at the bottom of the picture. After doing so, the subject must then decide where the dot will appear in the circle, among five alternative answers.

Development of General Coaching Strategy

To assess the degree to which the tests were susceptible to more traditional multiple-choice coaching, we scanned several popular coaching references (e.g., Barron's Educational Series, 1989; C.E.E.B., 1983; Steinberg, 1987) to develop a single-page handout listing hints on "Doing Better on Multiple-Choice Tests." These hints included such things as time management and guessing strategies. All subjects assigned to one of the general coaching conditions (see below) received the same handout, regardless of the test taken.

Subjects

Data were collected from 1,915 new Army recruits at Fort Jackson, South Carolina, in June of 1992. All subjects were

tested in two-hour sessions, in groups of 40 to 120 persons.⁷ Single testing sessions were conducted in the evenings on weekdays and three sessions were conducted on Saturdays.

Testing Schedules

Subjects were divided into one of fifteen groups, depending upon which of the three tests was involved, what kind of coaching, if any, was given, and whether or not subjects received coaching after or before practice.⁸ In all cases, retesting was on the same test form as the first testing. For each of the three tests, subjects were assigned to one of five conditions:

- 1) Specific Coaching After Practice. Subjects took one of the three tests, then listened to the audio tape and studied the workbook containing the specific coaching strategy for the test, then retook the test.
- 2) General Coaching After Practice. Subjects took one of the three tests, then received the handout giving coaching on general test-taking strategies before retaking the test (all subjects, regardless of the test taken, received the same handout).
- 3) Practice Only. Subjects took one of the three tests, then had a short break before re-taking the same test.
- 4) Specific Coaching Before Practice. Subjects received specific coaching before taking the test for the first time - after a short break, subjects retook the test.
- 5) General Coaching Before Practice. Subjects received general coaching before taking the test for the first time - after a short break, subjects retook the test.

Posttest Questionnaire

A posttest questionnaire was designed to elicit various types of data. Several versions of the questionnaire were developed to correspond to each of the tests and coaching conditions (specific, general, and practice). For purposes of the present research, four types of information were collected: 1) Ability to Recall Coaching Strategies, 2) Self-reported Use of Strategies, 3) Perceived Usefulness of Strategies, and 4) Perceived Likelihood of Coaching Like Ours if Spatial Testing

⁷In general, groups receiving specific coaching were small, while those receiving general coaching or practice only were larger.

⁸The second column in Appendixes A, B, and C shows the testing procedures for each of the fifteen groups.

Became Operational. All subjects completed the questionnaire after the main experiment.

Results

Within-Subjects Analyses

Table 1 shows within-subjects data for Number Correct, Speed, and Accuracy scores. [For these analyses, Speed was defined as the number of items attempted and Accuracy was the proportion of attempted items that were gotten correct.]⁹ Most effect sizes were positive; mean scores on the second testing in most groups were higher than means on the first testing. These effect sizes are, in general, very similar to others found for spatial tests (cf. Silva & Busciglio, 1993). In addition to this overall trend, several more specific results are also noteworthy:¹⁰

- Among the groups receiving Specific Coaching After Practice, Orientation had the largest effect size and Assembling Objects had the smallest. Among the groups receiving Practice Only, this order was reversed.
- As expected, effect sizes for General Coaching After Practice were very similar across the three tests and generally were smaller than those for Specific Coaching.
- Somewhat less expected was the finding that effect sizes for General Coaching After Practice were also smaller than those for Practice Only, on all tests except Orientation.
- For groups receiving coaching before practice, general coaching usually led to larger effect sizes than did specific coaching.
- In every case, improvements in Speed scores were highest for Assembling Objects and lowest for Orientation; for gains in Accuracy, this order was almost entirely reversed.¹¹

⁹Readers will note that the Number Correct score is therefore a multiplicative function of both Speed and Accuracy.

¹⁰More complete statistics on Number Correct, Speed, and Accuracy scores can be found in Appendices A, B, and C, respectively.

¹¹Readers might be interested in several trends revealed in Appendixes A, B, and C. First of all, improvements in speed, but not accuracy, seem to have been constrained somewhat by a ceiling effect, as shown by: a) mean posttest scores that were very close to the maximum possible (i.e., 36 on Assembling Objects, 30 on Figural Reasoning, 24 on Orientation), and b) smaller standard deviations of posttest, as opposed to pretest, scores. Also, males and females differed markedly on why their number correct scores increased as a result of Specific Coaching After Practice on the Assembling Objects test. As Appendixes B and C show, males achieved greater speed while females became more accurate.

Table 1

Effect Sizes and Significance Tests for Within-Subjects Analysis of Coaching and Practice Effects

Testing Schedule	Test	Type of Score					
		No. Correct		Speed		Accuracy	
		Eff	t	Eff	t	Eff	t
<u>Specific Coaching After Practice:</u>	AO	.32	4.4***	.43	6.5***	.04	0.5
	FR	.57	9.8***	.40	6.0***	.37	6.0***
	OR	.90	13.7***	.32	4.8***	.82	13.0***
<u>General Coaching After Practice:</u>	AO	.33	3.6***	.48	5.1***	.05	0.8
	FR	.31	5.6***	.45	5.3***	.07	1.0
	OR	.29	4.1***	.30	2.9**	.22	3.1**
<u>Practice Only:</u>	AO	.66	8.6***	.79	10.3***	.08	1.2
	FR	.49	5.6***	.57	5.8***	.12	1.4
	OR	.15	2.1*	-.39	-1.4	.17	2.3*
<u>Specific Coaching Before Practice:</u>	AO	.46	6.7***	.74	11.1***	-.06	-1.1
	FR	.04	0.7	.39	5.5***	-.15	-3.0**
	OR	.18	4.2***	.20	2.8**	.16	3.7***
<u>General Coaching Before Practice:</u>	AO	.38	5.4***	.59	7.8***	-.01	-0.2
	FR	.24	2.9**	.36	3.8***	.08	1.1
	OR	.32	3.3**	.14	1.0	.31	3.2**

Note. Eff = Effect size = (2nd Test mean - 1st Test mean)/SD on 1st Test. t = within-subjects t-test. ***p<.001. **p<.01. *p<.05.

Between-Subjects Analyses

Along with the within-subjects analyses reported above, we assessed several between-subjects effects. That is, we wanted to test for group effects in "score gains" between the first and second testings. For this purpose, we employed Analysis of Covariance (ANCOVA), using the first testing as the covariate and the second testing, as adjusted for the covariate, as the dependent variable.¹²

¹²Traditionally, this kind of analysis has been done in three types of designs: 1) Repeated measures ANOVA, 2) Oneway ANOVA with gain scores, and 3) Analysis of Covariance (ANCOVA). Although there remains debate as to which design is best, it is generally held that the ANCOVA design is superior. More specifically, it has been shown that the first two designs are algebraically equivalent and that both are equivalent to the ANCOVA when the pretest and posttest are perfectly correlated. However, when this is not the case - as in the present analyses, the

Effects of Coaching After Practice. This analysis assessed the extent to which specific and general coaching after practice led to significantly greater score gains than did practice only. Table 2 shows the results of the ANCOVA and subsequent comparison of [ANCOVA adjusted] cell means for all three tests.

The top portion of Table 2 shows results for the Number Correct scores. The Orientation test was the only measure for which specific coaching produced score gains that were significantly greater than those for practice only. For all three tests general coaching led to score gains that were about equal to or significantly smaller than those for practice alone. Specific coaching was more effective than general coaching for all tests except Assembling Objects.

The middle portion of Table 2 shows only one significant group effect on gains in speed scores (i.e., number of items attempted). Namely, that specific coaching on the Assembling Objects test led to a significantly lower gain than did either general coaching or practice alone.

Finally, the bottom portion of Table 2 shows group effects for gains in accuracy (proportion of attempted items that were gotten correct). For the Figural Reasoning and Orientation tests, specific coaching led to greater gains in accuracy than did either general coaching or practice.

Effects of Coaching Before Practice. The final between-subjects analyses compared the effects of general and specific coaching before practice to those of practice only. We once again used ANCOVA.

The results, as displayed in Table 3, contain very few significant effects. The top portion of the table shows that coaching, if anything, led to lower gains in Number Correct than did practice only; this is the case for general coaching on the Assembling Objects test and both types of coaching on the Figural Reasoning measure. The middle portion of the table shows no significant group effects for gains in speed. As the bottom portion shows, the only significant group difference for gains in accuracy was a lower gain for individuals receiving specific coaching on the Figural Reasoning test.

ANCOVA is generally the more precise, and thus preferable (cf. Cook & Campbell, 1979). A previous analysis of some of our results (Busciglio, 1992) used the gain score design. Although some specific comparisons led to slightly different results, the overall conclusions reached in that earlier paper are the same as those reported here.

Table 2

Between-Subjects Analysis of Effects of Coaching After Practice and Practice Only

		Test					
		Assembling Objects		Figural Reasoning		Orientation	
<u>Number Correct:</u>							
ANCOVA	F	df	F	df	F	df	
	2.67	2,391	8.93***	2,315	47.60***	2,431	
<u>Adjusted Cell Means</u>							
	N	Mean	N	Mean	N	Mean	
Specific	167	23.2 ab	163	22.6 a	222	15.8 a	
General	108	22.6 a	96	20.9 b	111	11.3 b	
Practice	120	24.5 b	60	22.2 a	102	11.0 b	
<u>Speed:</u>							
ANCOVA	F	df	F	df	F	df	
	4.47*	2,391	0.13	2,315	0.88	2,431	
<u>Adjusted Cell Means</u>							
	N	Mean	N	Mean	N	Mean	
Specific	167	32.7 a	163	29.5 a	222	23.6 a	
General	108	33.8 b	96	29.6 a	111	23.8 a	
Practice	120	34.3 b	60	29.6 a	102	23.6 a	
<u>Accuracy:</u>							
ANCOVA	F	df	F	df	F	df	
	1.04	2,391	12.56***	2,315	45.19***	2,431	
<u>Adjusted Cell Means</u>							
	N	Mean	N	Mean	N	Mean	
Specific	167	.712 a	163	.772 a	222	.662 a	
General	108	.684 a	96	.702 b	111	.475 b	
Practice	120	.704 a	60	.737 b	102	.477 b	

Note. Means are not significantly different ($p < .05$) from others in the same column marked with the same letter (a,b).
 * $p < .05$. ** $p < .01$. *** $p < .0001$.

Table 3

Between-Subjects Analysis of Effects of Coaching Before Practice and Practice Only

		Test					
		Assembling Objects		Figural Reasoning		Orientation	
<u>Number Correct:</u>							
ANCOVA	F	df	F	df	F	df	
	3.18*	2,388	7.85***	2,337	1.20	2,309	
Adjusted Cell Means	N	Mean	N	Mean	N	Mean	
Specific	155	25.7 a	171	20.9 a	152	13.6 a	
General	117	24.1 b	110	21.6 a	59	13.8 a	
Practice	120	25.6 a	60	23.2 b	102	12.9 a	
<u>Speed:</u>							
ANCOVA	F	df	F	df	F	df	
	1.08	2,388	0.54	2,337	1.10	2,309	
Adjusted Cell Means	N	Mean	N	Mean	N	Mean	
Specific	155	34.9 a	171	29.7 a	152	23.7 a	
General	117	34.5 a	110	29.8 a	59	23.9 a	
Practice	120	34.3 a	60	29.6 a	102	23.6 a	
<u>Accuracy:</u>							
ANCOVA	F	df	F	df	F	df	
	1.16	2,388	4.90**	2,337	0.77	2,309	
Adjusted Cell Means	N	Mean	N	Mean	N	Mean	
Specific	155	.735 a	171	.703 a	152	.564 a	
General	117	.710 a	110	.738 b	59	.583 a	
Practice	120	.731 a	60	.749 b	102	.548 a	

Note. Means are not significantly different ($p < .05$) from others in the same column marked with the same letter (a,b).

* $p < .05$. ** $p < .01$. *** $p < .0001$.

Analysis of Posttest Questionnaire Responses

After scoring the questionnaires, cleaning the data, and matching results with the data base of test scores from the larger coaching experiment, we obtained 1,894 usable questionnaires. Results are reported for each question separately below.

1. Ability to Recall Coaching Strategies. In the specific coaching version of the questionnaire, subjects were asked to list "the steps we taught you" for eliminating wrong answers and/or recognizing the correct answers on the tests. There were six such steps in the specific coaching for the Assembling Objects and Figural Reasoning tests, and four for the Orientation test. In the general coaching version, the same for all three tests, respondents were asked to list the six "ways we taught you for getting better scores on multiple-choice tests." We scored all versions of the questionnaire by giving subjects a point for each step (or strategy) they could remember.

Table 4 shows the relationship of gain scores to number of coaching strategies remembered.¹³ On all three tests, subjects on average remembered more than half the specific coaching strategies. Subjects in the general coaching groups could recall an average of one to three of the six steps. Surprisingly, all the correlations between number of strategies remembered and gain scores are small and none attained statistical significance.

2. Self-reported Use of Strategies. Subjects in the coaching groups were asked to indicate the extent to which they used the coaching strategies during the experiment. Responses included: (1) I used the coaching strategies as taught, (2) I used part(s) of the strategies, and (3) I did not use the strategies, or I tried the strategies and stopped.¹⁴

Table 5 shows the relationship of gain scores to subjects' self-reported use of the coaching strategies. There is some evidence that subjects using more of the specific and general strategies had higher gain scores. For example, on the Figural Reasoning test, receptees using all of the specific coaching strategies had significantly higher gain scores than did those

¹³For ease of interpretation, we used gain scores for a number of the following analyses. Responses in Tables 4, 5, and 6 were from subjects in the pretest-posttest (i.e., coaching after practice) groups. Responses in Table 7 were from all coached subjects.

¹⁴Subjects who did not use the strategies, or stopped using them, were asked why. Response options were: (4) I forgot the strategies, (5) I was not sure that I understood the strategies, (6) I thought my way was better, (7) I thought the strategies took too long, and (8) other, please specify. Our analyses of these data did not uncover any striking results, and will not be included in this paper.

Table 4

Relationship of Gain Scores to Number of Coaching Strategies Remembered

<u>Test</u> Type of Coaching	Average Number Remembered ^a	Correlation With Gain Scores
<u>Assembling Objects</u>		
Specific	3.39	.07
General	1.65	.16
<u>Figural Reasoning</u>		
Specific	5.12	-.07
General	1.72	.06
<u>Orientation</u>		
Specific	2.48	.05
General	2.24	-.04

Note. ^aThe maximum was 6 for all coaching except specific coaching on the Orientation test, where it was 4.

using only part or none of the strategies.¹⁵ On the Orientation test, those using all or part of the specific coaching strategies had significantly higher gain scores than those not using the strategies. For the general coaching groups, significant results were obtained for the Assembling Objects and Figural Reasoning groups.

3. Perceived Usefulness of Strategies. Subjects receiving specific or general coaching were asked, "How much do you think your test score improved as a result of the coaching you received in this session?" Responses were on a four-point Likert scale from "a great deal" to "not at all."

Table 6 shows the relationship of gain scores to receptees' perceptions of the improvement of test scores due to coaching. Although patterns of mean differences vary with type of coaching and test, in every case of specific coaching, receptees answering "a great deal" had significantly larger gain scores than did those indicating "not at all." This same pattern obtained for subjects receiving general coaching on the Assembling Objects test.

4. Perceived Likelihood of Coaching Like Ours if Spatial Testing Became Operational. Subjects in the specific and general coaching groups were asked "If spatial tests like those you just

¹⁵Results for the Assembling Objects test show a similar trend, but were not significant.

Table 5

Relationship of Gain Scores to Self-Reported Use of Coaching Strategies

<u>Test</u> Level of Use	<u>Specific Coaching</u>		<u>General Coaching</u>	
	N	Mean Gain Score	N	Mean Gain Score
<u>Assembling Objects</u>				
All	72	3.10 a	37	5.43 a
Part	45	1.69 a	27	4.00 ab
None	43	0.61 a	39	-0.21 b
Overall F		2.12		5.52**
<u>Figural Reasoning</u>				
All	102	3.77 a	42	2.81 a
Part	28	1.54 b	27	0.67 b
None	30	1.10 b	24	1.79 ab
Overall F		8.65**		3.73*
<u>Orientation</u>				
All	116	6.98 a	44	1.05 a
Part	35	5.66 a	32	2.16 a
None	43	1.88 b	33	1.15 a
Overall F		13.21***		0.87

Note. Means do not differ significantly from others in the same column with the same letter (a,b,c), by Tukey HSD test. *** $p < .0001$. ** $p < .01$. * $p < .05$.

took were made a requirement for getting into the Army, how likely would it be for someone to coach people on the spatial tests in a manner similar to the coaching you received in this session?" Table 7 shows the results. As can be seen, only a very small proportion of receptees stated that this was "not likely at all."

Discussion

Overall, our within-subjects results (Table 1) show that the Project A spatial tests involved in this research are subject to significant, nontrivial coaching and practice effects that are similar to those found in previous research on spatial tests (cf. Silva & Busciglio, 1993). In general, these effects are due to increases in both speed and accuracy.

Table 6

Relationship of Gain Scores to Recipients' Perceptions of the Improvement of Test Scores Due to Coaching

<u>Test</u> Response Category	<u>Specific Coaching</u>		<u>General Coaching</u>	
	N	Mean Gain Score	N	Mean Gain Score
<u>Assembling Objects</u>				
A Great Deal	16	5.81 a	4	10.50 a
A Moderate Amount	50	3.92 ab	38	4.92 ab
Only a Little	57	1.37 cb	35	2.83 ab
Not at All	33	-1.70 c	22	-2.32 b
Overall F		8.07***		6.23**
<u>Figural Reasoning</u>				
A Great Deal	23	5.35 a	2	3.00 a
A Moderate Amount	51	3.98 ab	25	2.80 a
Only a Little	65	2.12 cb	49	1.67 a
Not at All	21	0.24 c	17	1.18 a
Overall F		10.16***		1.05
<u>Orientation</u>				
A Great Deal	72	8.72 a	9	1.78 a
A Moderate Amount	42	7.07 a	25	2.80 a
Only a Little	48	2.08 b	42	1.33 a
Not at All	29	1.41 b	29	0.66 a
Overall F		24.73***		1.32

Note. Means do not differ significantly from others in the same column with the same letter (a,b,c), by Tukey HSD test. *** $p < .0001$. ** $p < .01$. * $p < .05$.

Perhaps of greater interest are the between-subjects analyses shown in Tables 2 and 3. First, there is no evidence that our general coaching was any more effective than practice alone. Several explanations are possible:

- Nonspatial coaching may simply be inappropriate for spatial tests.
- Overt hints about guessing may have led subjects to spend less time attempting to work items before guessing.
- The media used may have been too passive; it may have been preferable to have some type of "live" presentation of the material, such as the audio-tape and workbook mode used with the specific coaching material.

Table 7

Perceived Likelihood of Coaching Like Ours if Spatial Testing Became Operational

<u>Test</u> Response Category	<u>Specific Coaching</u>		<u>General Coaching</u>	
	No.	Pct.	No.	Pct.
<u>Assembling Objects</u>				
Extremely Likely	17	24.6	14	21.2
Very Likely	22	31.9	19	28.8
Somewhat Likely	24	34.8	27	40.9
Not Likely at All	6	8.7	6	9.1
<u>Figural Reasoning</u>				
Extremely Likely	23	27.1	8	13.8
Very Likely	29	34.1	17	29.3
Somewhat Likely	27	31.7	29	50.0
Not Likely at All	6	7.1	4	6.9
<u>Orientation</u>				
Extremely Likely	21	30.9	12	16.2
Very Likely	21	30.9	23	31.1
Somewhat Likely	21	30.9	28	37.8
Not Likely at All	5	7.3	11	14.9

In some comparisons (e.g., gains in Number Correct scores due to Coaching After Practice on the Assembling Objects and Figural Reasoning tests - see top portion of Table 2) the general coaching groups actually did less well than the practice only groups. In these cases, the coaching intervention may have had a negative, or inhibitory impact on practice effects.

Turning to specific coaching, the top portion of Table 2 shows that, for the Assembling Objects and Figural Reasoning tests, Specific Coaching After Practice resulted in about the same pretest-posttest score gains as did Practice Only.¹⁶ In retrospect, we are not surprised by this finding, given the somewhat greater-than-desired length and complexity of the coaching strategies designed for these measures. Despite our best attempts, we were unable to discover any simple, comprehensive "tricks" that would be equally effective, or nearly so, across all items of these tests. Because of this, our

¹⁶Readers should note that a ceiling effect is not a likely explanation, since pretest means for all groups are not close to the maximum possible. Also, there is not a great difference in the standard deviation of pretest scores between the groups.

coaching may have been, for some subjects, little more than concentrated practice. Given that we spent a great deal of time and effort in an attempt to develop the shortest, easiest strategies possible, we doubt that any effective coaching on these tests can take place quickly and easily. Thus, for these two tests, coaching books like those available for ASVAB would probably be needed for any great improvement in scores.¹⁷

For the Orientation test, on the other hand, Specific Coaching After Practice led to much larger pretest-posttest gains than did Practice Only. Once again, we are not surprised. Unlike that for the Assembling Objects and Figural Reasoning tests, coaching for the Orientation test involved teaching a simple, straight-forward "trick," or hint, that could be used with all items. That this is a true coaching effect also seems in line with the fact that it came about mostly by a very large increase in accuracy, as opposed to speed. We would strongly recommend, therefore, that attempts be made to reduce the effectiveness of such coaching before this measure is used as part of the Army's testing procedures.¹⁸

The analysis shown in Table 3 was done to determine the effects of Coaching Before Practice. Such coaching generally led to score gains that were equal to or lower than those for Practice Only. Perhaps this occurred because the coached subjects were approaching the top of their learning curves and thus had less room for improvement than did their noncoached counterparts.¹⁹

Concerning the sizeable effects of Practice Only on the Assembling Objects and Figural Reasoning tests, several potential countermeasures might be explored, such as:²⁰

- 1) giving all examinees more practice items to complete immediately before the test itself;²¹

¹⁷This, of course, speaks to the quality of these measures, since many types of "quick-and-easy" coaching strategies are meant to exploit flaws in test design, flaws that are missing from such well-developed instruments as the ASVAB.

¹⁸A concern here is that countermeasures that change the actual content of test items may alter the nature of the abilities being measured.

¹⁹Unlike a ceiling effect, this conjecture cannot be supported or refuted by any of our data.

²⁰However, a comparison of our results with those of Peterson (1987), as cited earlier in this paper, suggests that our practice effect sizes may rapidly deteriorate over a two week period.

²¹However, as Silva and Busciglio (1993) have pointed out, even when practice and/or coaching are available to all members of a group, their effects are still problematic, given their unknown impact upon the construct validity of tests.

- 2) including, in future ASVAB orientation materials, motivation and opportunities for individual practice before the test session;
- 3) expanding the range of content of test items, thus reducing the effectiveness of any short, simple practice strategy.²²

Finally, we wish to say a few words about our posttest questionnaire results. This instrument was designed, among other things, to "shed light" on any hard-to-understand results from the objective data. For example, a lack of coaching effects might be explained by a low rate of self-reported use of coaching. We believe, however, that there were very few surprising results in either the objective data or in the questionnaire responses. Although correlations between gain scores and number of strategies remembered were somewhat low, there were generally strong relationships between gain scores and self-reported use of coaching. Also, subjects seemed to be aware of how much the coaching improved their scores. Finally, a very large proportion of receptees felt that it was at least somewhat likely that similar coaching strategies would be used in the future if the spatial tests became operational. We see the latter two findings as evidence for the reasonableness of the strategies we devised.

²²A number of comments should be made about this option. First, any such expansion of test content may threaten the construct validity of the measure. Secondly, practice effects may generalize across at least some dimensions of content variation. Finally, since item level analyses gave no clear indication that coaching or practice effects differed across items, or item types, it seems unlikely that such effects can be reduced by excluding certain types of items.

References

- Anastasi, A. (1982). Psychological testing (fifth edition). New York: MacMillan.
- Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. Sex Roles, 20, 327-344.
- Barron's Educational Series (1989). How to prepare for the Armed Forces Test (3rd. Ed. pp. 7-9). New York: Barron's Educational Series, Inc.
- Brinkmann, E. H. (1966). Programmed instruction as a technique improving spatial visualization. Journal of Applied Psychology, 50, 179-184.
- Busciglio, H.H. (1992, October). Assessing the coachability of Project A spatial tests: Empirical results. Paper presented at the annual meeting of the Military Testing Association, San Diego.
- Campbell, J.P., & Zook, L.M. (Eds.). (1991). Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A. ARI Research Report 1597. Alexandria, VA: U.S. Army Research Institute.
- C.E.E.B. (1983). 10 SATs: Scholastic aptitude tests of the college board (p. 11). New York: College Entrance Examination Board.
- Conner, J. M., Schackman, M., & Serbin, L. A. (1978). Sex-related differences in response to practice on a visual-spatial test and generalization to a related test. Child Development, 49, 24-29.
- Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand McNally.
- Eysenck, H. J. (1962). Know your own I.Q.. Baltimore: Penguin.
- Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. Educational Communication and Technology, 33(4), 263-275.
- Gibson, E. J. (1953). Improvement in perceptual judgements as a function of controlled practice or training. Psychological Bulletin, 50, 401-431.
- Goldstein, A. G., & Chance, J. E. (1965). Effects of practice on sex-related differences in performance on Embedded Figures. Psychonomic Science, 3, 361-362.
- Jacobs, P. I. (1977). Up the IQ. New York: Wyden.
- Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. Journal of Educational Psychology, 76, 130-145.
- McGee, M. G. (1978). Effects of training and practice on sex differences in mental rotation test scores. Journal of Psychology, 100, 87-90.
- Peterson, N.G. (Ed.). (1987). Development and field test of the trial battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute.
- Saunderson, A. (1973). The effect of a special training programme on spatial ability test performance. New Guinea Psychologist, 5, 15-23.
- Sherman, J. A. (1974). Field articulation, sex, spatial visualization, dependency, practice, laterality of the brain and birth order. Perceptual and Motor Skills, 38, 1223-1235.
- Silva, J.M., & Busciglio, H.H. (1993, January). Impact of Practice and Coaching. Paper presented at the Military Accession Policy Working Group Meeting, San Diego.
- Steinberg, E.P. (1987). Scoring high on the Armed Forces test (pp. 27-30). New York: Pocket Books.
- Stericker, A., & LeVesconte, S. (1982). Effect of brief training on sex-related differences in visual-spatial skill. Journal of Personality and Social Psychology, 43, 1018-1029.

Appendix A

Statistics For Number Correct Scores

Test Schedule ^a		Sex	N	1st Test		2nd Test		Effect Size ^b	t ^c	
				M	SD	M	SD			
<u>Specific Coaching After Practice:</u>										
AO	O AO O	M	117	23.2	6.9	25.0	7.8	0.27	2.958**	
		F	50	20.9	6.7	24.0	6.9	0.46	3.845***	
		All	167	22.5	6.9	24.7	7.5	0.32	4.434***	
FR	O FR O	M	105	20.3	4.9	23.2	4.4	0.60	7.455***	
		F	58	20.2	5.5	23.1	4.2	0.53	6.566***	
		All	163	20.3	5.1	23.2	4.3	0.57	9.817***	
OR	O OR O	M	222	10.8	6.2	16.3	6.7	0.90	13.657***	
<u>General Coaching After Practice:</u>										
AO	O GN O	M	108	19.1	8.4	21.9	8.8	0.33	3.648***	
FR	O GN O	F	96	18.0	6.2	19.9	6.0	0.31	5.643***	
OR	O GN O	F	111	8.7	5.5	10.3	6.0	0.29	4.126***	
<u>Practice Only:</u>										
AO	O O	M	120	18.2	7.5	23.1	8.4	0.66	8.594***	
FR	O O	F	60	19.6	5.5	22.3	4.9	0.49	5.625***	
OR	O O	M	55	11.3	6.4	12.6	7.4	0.21	2.248*	
		F	47	8.7	5.3	9.1	5.8	0.08	0.680	
		All	102	10.1	6.0	11.0	6.9	0.15	2.135*	
<u>Specific Coaching Before Practice:</u>										
AO	AO O O	M	51	25.5	7.0	28.1	6.6	0.37	2.951**	
		F	104	24.7	6.7	28.1	5.9	0.51	6.206***	
		All	155	24.9	6.8	28.1	6.1	0.46	6.721***	
FR	FR O O	M	171	21.2	5.5	21.4	6.4	0.04	0.712	
OR	OR O O	M	116	14.5	7.6	15.8	8.0	0.17	3.345**	
		F	36	11.7	7.1	13.3	7.7	0.23	2.750**	
		All	152	13.9	7.6	15.2	8.0	0.18	4.209***	

(Continued)

Appendix A (Continued)

Statistics For Number Correct Scores

Test Schedule ^a	Sex	N	1st Test		2nd Test		Effect Size ^b	t ^c
			M	SD	M	SD		
<u>General Coaching Before Practice:</u>								
AO	GN O O	M	117	20.7 7.2	23.5 8.3	0.38	5.390***	
FR	GN O O	M	110	20.2 4.5	21.3 5.3	0.24	2.881**	
OR	GN O O	M	59	10.8 5.8	12.7 7.6	0.32	3.250**	

^aO = Testing; GN = General Coaching; AO, FR, OR = Specific coaching on Assembling Objects, Figural Reasoning, and Orientation tests, respectively. ^bEffect size = (2nd Test mean - 1st Test mean)/SD on 1st Test. ^ct is for dependent groups test: ***p<.001 **p<.01 *p<.05.

Appendix B

Statistics For Speed Scores

Test Schedule ^a	Sex	N	1st Test		2nd Test		Effect Size ^b	t ^c
			M	SD	M	SD		
<u>Specific Coaching After Practice:</u>								
AO O AO O	M	117	30.8	6.1	34.1	3.9	.54	6.63***
	F	50	28.6	7.0	30.1	7.2	.21	1.91
	All	167	30.1	6.5	32.9	5.4	.43	6.53***
FR O FR O	M	105	27.8	4.1	29.5	1.7	.42	5.16***
	F	58	28.7	3.0	29.8	1.2	.36	3.17**
	All	163	28.1	3.8	29.6	1.6	.40	6.04***
OR O OR O	M	222	22.5	3.5	23.6	1.6	.32	4.84***
<u>General Coaching After Practice:</u>								
AO O GN O	M	109	30.8	7.5	34.3	5.9	.48	5.09***
FR O GN O	F	96	27.3	4.8	29.5	2.5	.45	5.30***
OR O GN O	F	111	22.9	3.1	23.8	1.2	.30	2.93**
<u>Practice Only:</u>								
AO O O	M	120	27.5	7.6	33.5	5.1	.79	10.28***
FR O O	F	60	26.4	5.0	29.3	2.4	.57	5.83***
OR O O	M	55	23.8	0.5	23.9	0.4	.11	0.68
	F	47	23.9	0.5	23.4	2.0	-.99	-1.65
	All	102	23.8	0.5	23.6	1.4	-.39	-1.35
<u>Specific Coaching Before Practice:</u>								
AO AO O O	M	51	31.9	5.2	35.5	1.7	.68	5.77***
	F	104	30.3	6.1	35.0	2.7	.78	9.48***
	All	155	30.8	5.9	35.2	2.4	.74	11.06***
FR FR O O	M	171	28.4	3.4	29.8	1.4	.39	5.46***
OR OR O O	M	116	23.7	1.9	23.9	0.3	.15	1.56
	F	37	20.8	5.2	22.6	2.4	.36	2.32*
	All	153	23.0	3.3	23.6	1.3	.20	2.75**

(Continued)

Appendix B (Continued)

Statistics For Speed Scores

Test Schedule ^a	Sex	N	1st Test		2nd Test		Effect Size ^b	t ^c	
			M	SD	M	SD			
<u>General Coaching Before Practice:</u>									
AO	GN O O	M	117	31.2	6.3	34.9	4.4	.59	7.80***
FR	GN O O	M	110	29.1	2.5	30.0	0.1	.36	3.76***
OR	GN O O	M	59	23.8	1.1	23.9	0.3	.14	1.03

Note. Speed = number of items attempted. ^aO = Testing; GN = General Coaching; AO, FR, OR = Specific coaching on Assembling Objects, Figural Reasoning, and Orientation tests, respectively. ^bEffect size = (2nd Test mean - 1st Test mean)/SD on 1st Test. ^ct is for dependent groups test: ***p<.001 **p<.01 *p<.05.

Appendix C

Statistics For Accuracy Scores

Test Schedule ^a	Sex	N	1st Test		2nd Test		Effect Size ^b	t ^c	
			M	SD	M	SD			
<u>Specific Coaching After Practice:</u>									
AO O AO O	M	117	.755	.18	.734	.22	-.12	-1.50	
	F	50	.732	.15	.801	.13	.47	3.77***	
	All	167	.748	.17	.754	.20	.04	0.51	
FR O FR O	M	105	.736	.16	.788	.14	.33	3.86***	
	F	58	.701	.18	.775	.13	.42	5.36***	
	All	163	.723	.16	.783	.14	.37	5.99***	
OR O OR O	M	222	.477	.26	.689	.27	.82	13.04***	
<u>General Coaching After Practice:</u>									
AO O GN O	M	108	.615	.24	.633	.23	.05	0.75	
FR O GN O	F	96	.655	.20	.668	.19	.07	1.03	
OR O GN O	F	111	.382	.23	.434	.25	.22	3.10**	
<u>Practice Only:</u>									
AO O O	M	120	.673	.22	.691	.22	.08	1.16	
FR O O	F	60	.740	.16	.760	.15	.12	1.40	
OR O O	M	55	.472	.26	.529	.31	.21	2.20*	
	F	47	.365	.22	.389	.24	.11	0.93	
	All	102	.423	.25	.464	.29	.17	2.28*	
<u>Specific Coaching Before Practice:</u>									
AO AO O O	M	51	.796	.17	.792	.19	-.02	-0.17	
	F	104	.815	.15	.802	.15	-.09	-1.45	
	All	155	.809	.16	.799	.16	-.06	-1.06	
FR FR O O	M	171	.746	.18	.718	.21	-.15	-3.03**	
OR OR O O	M	116	.611	.31	.660	.33	.16	3.14**	
	F	36	.530	.27	.567	.31	.17	1.88	
	All	152	.592	.30	.638	.33	.16	3.66***	

(Continued)

Appendix C (Continued)

Statistics For Accuracy Scores

Test Schedule ^a	Sex	N	<u>1st Test</u>		<u>2nd Test</u>		Effect Size ^b	t ^c	
			M	SD	M	SD			
<u>General Coaching Before Practice:</u>									
AO	GN O O	M	117	.669	.20	.666	.23	-.01	-0.19
FR	GN O O	M	110	.697	.16	.709	.18	.08	1.07
OR	GN O O	M	59	.456	.24	.530	.31	.31	3.21**

Note. Accuracy = Proportion of attempted items gotten correct. ^aO = Testing; GN = General Coaching; AO, FR, OR = Specific coaching on Assembling Objects, Figural Reasoning, and Orientation tests, respectively. ^bEffect size = (2nd Test mean - 1st Test mean)/SD on 1st Test. ^ct is for dependent groups test: ***p<.001 **p<.01 *p<.05.